DOCUMENT RESUME

ED 027 922 LI 001 352

By-Brandhorst, W. T.

Data Form and Availability and the Design of Computerized Retrieval Systems Dealing with Bibliographic Entities.

Pub Date 69

Note-39p.; Paper presented at the Annual Midwest Conference, American Statistical Association (16th, Chicago, March 28, 1969)

EDRS Price MF-$0.25 HC-$2.05

Descriptors-*Automation, *Bibliographic Citations, Coordination, Data Analysis, Electronic Data Processing, Information Processing, Information Retrieval, *Information Storage, *Information Systems, Input Output, Libraries, Library Materials, Programing, Standards, *Systems Development

An analysis of existing computerized data banks in science and technology reveals that nearly half of them involve the storage and retrieval of bibliographic data. Activity in this area has been independent and autonomous. This situation is now giving way to a new environment which involves cooperation, standards, and a rigorous rational analysis of the traditional raw materials and processes of librarianship. There is a genuine rapprochement between librarians and computer specialists, resulting in a scientific approach to the problems posed by the control and retrieval of bibliographic entities. It is observed that the design of systems dealing with bibliographic data must respond to pressures generated by the structure of the data itself, as well as to inter-system and user requirements. These pressures can take effect on several levels, ranging from programming subroutines, to file structure, to the organization of access modes and output formats. Examples of each are provided. This is a critical period for library automation, for the standardization of machine input records, and for the design of retrieval systems dealing with these records. Decisions made today will affect the access to bibliographic data for decades to come. (Author/JB)

TITLE: Data Form and Availability and the Design of Computerized Retrieval Systems Dealing with Bibliographic Entities

AUTHOR: W. T. Brandhorst

AFFILIATION: Leasco Systems and Research Corporation

OCCASION: 16th Annual Midwest Conference, American Statistical Association

THEME: Management of Information for Decision. Statistics for Decision.

DATE: March 28, 1969

PLACE: Chicago, Illinois -- Pick Congress Hotel

CONTENTS

SUMMARY

An analysis of existing computerized data banks in science and technology reveals that nearly half of them involve the storage and retrieval of bibliographic data. Activity in this area has, in the past, been characterized by independent, autonomous efforts, each finding its own solutions to much the same set of problems. This situation is giving way to a new environment in which we find cooperation, standards, and a rigorous rational analysis of the traditional raw materials and processes of librarianship. There is evident a genuine rapprochement between librarians and computer specialists, resulting in a scientific approach to the problems posed by the control and retrieval of bibliographic entities.

It is observed that the design of systems dealing with bibliographic data must respond to pressures generated by the structure of the data itself, as well as to inter-system and user requirements. These pressures can take effect on several levels, ranging from programming subroutines, to file structure, to the organization of access modes and output formats. Examples of each are provided.

This is an exciting and critical period for library automation, for the standardization of machine input records, and for the design of retrieval systems dealing with these records. Librarians and information scientists are at the brink of a new era. Decisions being made today will affect the way we will have access to our bibliographic heritage for decades to come.

DATA FORM AND AVAILABILITY AND THE DESIGN OF COMPUTERIZED RETRIEVAL SYSTEMS DEALING WITH BIBLIOGRAPHIC ENTITIES

I. Introduction

I have been asked to talk to you today about some of the factors involved in designing computerized systems dealing with bibliographic data. By bibliographic data I mean the various data elements that have historically been used to describe documentary entities (books, articles, technical reports, theses, proceedings, etc.), both as physical objects and as intellectual (information-bearing) objects. Common examples of such data elements are: the title, the author, the date of publication, the conference name, the total number of pages, and the subject index terms. These elements are, in general, quite familiar to most of us and I needn't inventory them all here. We see them everyday whenever we peruse a citation, a reference, an abstract journal entry, an index, a library accession list, a selective dissemination of information (or SDI) announcement, or a common garden-variety 3" x 5" library catalog card. I'm sure that few of us manage to avoid looking at bibliographic data entirely during the course of a given day.

I don't want to spend a lot of time justifying this topic, though I must admit when I finally saw the entire program my initial reaction was to do exactly that. I think, rather, I would simply like to assert that there exists an intimate connection between documentary records, containing scientific and technical information, and the management process, and let it go at that, with perhaps one example. This fairly safe contention has perhaps nowhere been more dramatically exemplified than in the National Aeronautics and Space Administration. James Webb, former NASA administrator, stated in his 1968 Diebold lecture at Harvard, on technological change and management, that "The essence of the job

NASA has done is not that a new body of knowledge and technology has been brought into being. Most of the basic knowledge and basic technology was already at hand. The essence of our job has been that or organizing and managing the use of available knowledge and technology in a purposeful and effective way." (Ref. 41, p. 23)

For the purposes of this paper, I like to think of the term "Available Knowledge" as referring in this context, strongly, if not entirely, to documentary entities. This hasty nod to the decision-making process will have to suffice, at least for the time being, by way of justification.

II. Bibliographic Data Processing -- General Comments on the State-of-the Art

We happen to be about now at a rather exciting crossroads in the history of the machine processing of bibliographic data. This is, of course, a relative kind of thing and I'm sure that some of you will find the developments I am about to describe less than dramatic.

In 1968, an interesting reference work entitled Directory of Computerized Information in Science and Technology (Ref. 12) was published. This book contains entries for nearly 300 information systems. An analysis of these systems reveals that approximately 50% are concerned with the storage and retrieval of bibliographic data. (The others involve such data as neutron cross-sections, cancer test results, etc. and "fact retrieval" as opposed to document retrieval.) Virtually all of these bibliographic systems are autonomous and independent efforts which were designed to satisfy their own system requirements but which had no particular concern for anything outside their individual frameworks.

One of the systems that is treated in this reference work is that at the NASA Scientific and Technical Information Facility, in College Park, Maryland, which, up until last November, was operated for NASA by Leasco Systems & Research Corporation, and where I served as an Assistant Director. Back in 1962, when this Facility was first established, we sat down with NASA and developed detailed specifications for every element of the desired system. For example, our experts in reprographic technology drew up complete technical specifications for the microforms to be prepared: overall dimensions, distance from edge of card to frames, distance between frames, thickness, acceptable curl, image resolution, everything down to the smell of the film! In another area, I sat down with my counterparts and we ran through the entire set of bibliographic data elements, or at least we ran through as many as anybody connected with the job could then conceive of. We decided which ones we were

going to collect and we decided various details about the appearance of the things collected. We designed the system's standard bibliographic citation, which puts most of these elements in relation to one another, and we designed various sub-versions of the citation for special purposes, such as the selective dissemination (SDI) system.

In short, we designed a computerized system for handling NASA's bibliographic data, and within three months the primary product, an abstract journal with indexes, prepared entirely by this system, was rolling off the presses. Later in the first year, as the file grew to a decent size, we began to do our first important retrieval work in selectively pulling material from the file in response to specific queries. Each year since, the uses of the master file of bibliographic data and the information products flowing from it, have increased in number and sophistication. Selective Dissemination of Information (SDI) or current awareness systems of several types were developed. Continuing or recurring bibliographies on topics of major interest were begun. Distribution of bibliographic data on magnetic tapes to a body of field users was entered into; perhaps the U.S. government's first such effort. On-line, real-time access to the data bank was initiated on an experimental basis. The Facility became a virtual document-processing factory with raw materials, in the form of government R & D reports, entering the hopper in profusion from one end and a multitude of products, representing different packagings of the information in these documents, emanating from the output end.

Looking back at this early design activity now, I can appreciate better the problems we had simply because we were early in the game. There were no government-wide or professional standards for cataloging technical report literature. Nobody had even made up a really complete list of the kinds of things you ran into in this work. And there were certainly no standards or

even recommendations for a machine file structure. What data elements should be captured? Of those captured, which require separate tagging? What kind of overall structure is best: a directory (sometimes called a relative image) structure, or an embedded identifier structure? If the latter course is followed, should these appear with or without explicit length data? Should we program in a higher level language or a machine-oriented assembly language? There were questions on all levels to be answered, by everyone from the cataloger to the reference librarian to the systems programmer.

Now roughly the same was true in some other areas. There were then, for instance, no government-wide or professional specifications for microfiche. These came along a bit later via the coordinating efforts of the Government's Committee on Scientific and Technical Information (familiarly known as COSATI), and implementing them was, relatively speaking, no problem. As I hardly need to tell you, however, the same can rarely be said when you get very far down the road with a software system with a lot of interrelated parts. It isn't so much the file format you've chosen that kills you. This can be converted, albeit usually with limitations. It is the software that surrounds and manipulates the file and produces the system's various outputs that provides the inertia. What you do at the beginning, right or wrong, you often have to live with for some time until the next massive re-design, re-programming, or "re-computering" (Machine Replacement) effort.

The present situation is much improved over what we at the NASA Facility found in 1962, and over what every other designer of the early years of this decade found, whenever they began.

It can be regretted that these improvements didn't appear earlier, so as to make possible a greater degree of compatibility among the 150 information systems, referred to in the Directory I cited,

most of which began in the 60's. However, this is a fruitless kind of hindsight as the pioneer systems themselves were probably necessary in order to clearly establish a need and a base of experience.

The improvement in the situation can, I think, be described under two headings: (1) Standards, and (2) A new, systematic, even, if you will, "scientific", approach to the problems.

A. Standards (e.g. COSATI, interagency cooperation, Project MARC, etc.)

Through the efforts of COSATI, for example, government-wide standards have been arrived at for products such as microfiche, and functions such as descriptive cataloging of technical reports.

Through the cooperation of NASA and the Department of Defense (DID) in the preparation of their respective thesauri of scientific and technical terminology, we are, in effect, moving toward a standard pattern in the vocabulary area also.

The Library of Congress' Project MARC (MARC stands for MAchine-Readable Cataloging) has led, with the MARC II format, to a national standard, with impressive official support, for a computerized record for the communication of monographic bibliographic data between one organization and another. The studies and investigations that led to MARC have, however, done even more than that really. They have led to a groundswell of new sensitivity and awareness in the profession (and by profession I mean here librarians and information scientists together) of the nature of the basic data we work with.

This is one of the really exciting things I see happening in the profession and leads directly to my second heading.

B. Scientific Approach

Quite clearly there is a new, systematic, even scientific, approach to problems of bibliographic data. I have, in my own mind, always

considered this as dating from the publication, in 1965, of Buckland's report to the Council on Library Resources, entitled The Recording of Library of Congress Bibliographical Data in Machine Form (Ref. 10). This report laid the problems on the line and told everyone frankly that "at the present time there is no firm basis or set of standards for the use of bibliographic data in machine form. Even the present manual uses of the bibliographic data are not well defined.... The crux of the problem, which affects the long term use of any data recorded, is that very little is known formally about how bibliographic information will be machine processed to accomplish various objectives. This results in an inadequate set of specifications controlling what data is to be recorded and in what form."

This report made a number of observations that later proved extremely fundamental in nature. For example, it pointed out that most bibliographic data performs multiple functions, overlapping, for instance, into both the areas of control and search. It classified the various kinds of coding or identification found, or possible, for bibliographic data, ranging from the fully explicit to the so implicit that the data is hidden in all practicality from even a full scan and complex program manipulation (e.g. the difference between a personal author and a corporate author). It also laid on a few basic requirements that, happily, librarians proceeded to pick up and run with over the next few years; for example:

"Before a standard machine-readable record for bibliographic data is agreed upon, the library world should consider what additional elements require to be distinguished in the bibliographic entry.... It appears that card catalog data needs to be recorded for long term purposes since the data being encoded today will be in use 10, 25, and 50 years from now. Because new uses of the data are apt to emerge in its life span, we had better think now

about what should be contained in the record to have the best possibility of satisfying the new requirements." (p. 30)

In this same tradition, the report entitled The Identification of Data Elements in Bibliographic Records (Ref. 15), done for the United States of America Standard Institute (USASI), Z-39 Committee, by Ann Curran, now of Inforonics, is an extremely important fundamental step in a scientific approach to the problem. Without regard for any one provincial point of view, or any one library or type of literature dealt with, it spreads all the elements out for the first time, like a bunch of potsherds awaiting the archaeologist's hand.

The realization is also beginning to sink in that we librarians, experts in description that we style ourselves, have, on yet another level, not adequately described the phenomena we deal with. We have not adequately described our bibliographic descriptions! This may sound exotic, but I assure you it is an absolute necessity when working with the design of today s machine systems. A good example is the MARC staff paper, "Fields of Information on Library of Congress Catalog Cards: Analysis of a Random Sample." (Ref. 4). This study was necessary because the MARC investigators found they didn't know enough about bibliographic descriptions. What is the frequency of appearance of the various elements in these descriptions? How frequently do personal authors appear? Do illustration statements appear? What is the distribution of multiple authors? of multiple illustration statements? What is the distribution of materials across foreign languages? What is the maximum, minimum, and average length of titles and what does the distribution curve look like? What about the appearance of special characters? How many can be identified and with what frequency for each?

Some of these questions the MARC staff tried to answer. They had to. You can't define a character set for a computer system without answering the

last two, for example. Others, MARC itself hasn't gotten to yet but other people have. For example, if you are engaged in loading random access storage equipment with inverted files or serial citation files of bibliographic data, or if you are working with an on-line CRT system that is going to be moving this data around a lot in interactive fashion, you are going to ask yourself numerous questions having to do with the distribution and lengths of data elements.

When you think about it a little, you can't help but feel that the profession has been remiss. Even though there may not have been any major use for this data prior to the advent of machine systems, the fact that the data was not available when people looked for it seems a failure to follow a systematic and scientific approach to one's area of responsibility; something that a zoologist, for instance, would not have been guilty of with respect to his animals.

Even though I can call librarians remiss, however, they have been no more so than numerous other disciplines being hit for the first time with the unsentimental demands of the new computer technology, and at the same time I think the library profession _is_ responding gratifyingly to the computer challenge. More and more librarians are learning about "systems analysis," "automation," "programming," etc. Old long-established practices are being re-examined. Things are being hauled out into the light that had become virtually scriptural in the library science curriculum, and are being forced to re-justify their existence -- if they can.

For example, Wesley Simonton, of the University of Minnesota in his article "The Computerized Catalog: Possible, Feasible, Desirable?" (Ref. 40) does perhaps the best of several jobs in subjecting the cherished concept of "Main Entry" to a set of searching questions which are very rapidly bringing it back down to earth and into mortal perspective. (On this subject see also Ref. 6, p. 26)

The hallowed American Library Association (ALA) and Library of Congress (LC) filing rules are likewise receiving detailed cross-examinations. William Nugent, of Inforonics, in his article entitled "The Mechanization of the Filing Rules for the Dictionary Catalogs of the Library of Congress" (Ref. 35), addresses himself to the question of just what it would take in a computer system and a machine-readable record to achieve exact correspondence with the present rules. Perhaps in some cases we won't want to pay the price. Filing rules seem almost certain to receive some modification as a result of computerization.

And yet at the same time I feel that the profession is proceeding with an appropriate dignified haste. Frederick Kilgour, then Associate Librarian at Yale University, in his paper "Symbol-Manipulative Programming for Bibliographic Data Processing on Small Computers" (Ref. 25) states:

> "Perhaps the cardinal principle of a bibliographic data processing system is that the machine must not be allowed to impose its characteristics on the data or the procedure. In the case of library procedures, long experience has accrued; indeed libraries are thousands of years old, while books have been printed for hundreds of years. Lessons learned empirically, decades and perhaps centuries ago, should not be discarded because of machine characteristics or because of difficulties in program planning or coding."

As a librarian who has worked long and hard with systems and programming people, I agree wholeheartedly with Mr. Kilgour's observation. It is very easy for the computer types to discount puzzling library "habits." As Buckland stated it, "Definition of the function or uses of bibliographic data needs to be made by experienced librarians. In those cases where these uses have been left to programmers or machine salesmen, difficulties have arisen

before many thousand items of information have been processed." (Ref. 10, p. 32). It is up to the library profession to rationalize its practices, in the sense of basing them on rational principles. What can't be rationalized can be done without.

III. Factors in the Design of Bibliographic Data Systems

When I consider the problem of designing a document retrieval system, I can envisage the bibliographic data that will be the concern of the system, exercising three kinds of technical pressures on the designer. I say technical because I would like to set aside the economic cost-effectiveness pressures for the purposes of this discussion.

The data exercise certain pressures on the programmer by virtue of their basic attributes. These pressures are distinctly different from the pressures that the data in an accounting system, or an inventory system, or any other kind of system, exercise.

Likewise, there are certain pressures on the system designer that are attributable to the unique system requirements of bibliographic data. These pressures also stem from the basic nature of bibliographic data but have an effect on a higher level than programming subroutines. They affect such system characteristics as record format.

Thirdly, there are pressures on the system designer that arise because of the way that users require access to, and outputs from, this particular kind of data. These are somewhat less fundamentally tied to data structure and are subject to change depending on the user population.

I would like to spend the remainder of this paper providing examples of each kind of pressure.

A. Programming

Papers dealing with the programming aspects of bibliographic data are something of a rarity. About the best thing I found in my search for references was a paper by Sally Alanen, a programmer formerly at Yale University, entitled "A Library of Subroutines for Bibliographic Data Processing." (Ref.1) She

herself emphasizes the sparseness of material at the very beginning of her paper, when she says:

> "Computer processing of bibliographic data differs from other data processing in the types of operations that are most frequently performed. If the operations that are most useful for bibliographic programming can be identified, then tools can be devised to perform them efficiently and to save programming steps...... Little has been published about specific programming problems of bibliographic data processing...... There has not yet appeared an analysis and classification of subroutines for bibliographic data processing. This paper will classify such a set of subroutines and describe some of its components."

The subroutines she goes on to cover range over the following headings: input/output, compression coding, tag generators, array searching, comparison subroutines, array transfers, data packing and unpacking, data mode conversion, filing, and generalized sort-merge programs.

The single example that can perhaps be described in the fewest number of words concerns input/output (I/O). Since it is a characteristic of bibliographic systems that I/O consumes a high percentage of total processing time, Miss Alanen advises that a subroutine should be written with efficiency as its primary goal. It would permit unformatted, variable length records to be read into and written out of buffers in core memory so that I/O requests can, wherever possible, be serviced individually from the buffers, which are subsequently refilled. Though the example may seem elementary, I think it demonstrates my point sufficiently so that more specialized subroutines needn't be described here. Numerous additional examples and associated discussion are contained in Miss Alanen's paper and I recommend it to those wishing to pursue the subject.

B. Systems Design

Coming up a notch from the programmer's problems with bibliographic

data, we can assume the systems analyst point of view. John Knapp's paper entitled "Design Considerations for the MARC Magnetic Tape Formats" (Ref. 27) is a good example of such a viewpoint. In this paper, Knapp, of LC's Information Systems Office, does an excellent job of demonstrating how librarians have in the past relied on the technique of formatting to communicate large amounts of information to the user. The 3 X 5 catalog card, arrayed in their millions in so many large research libraries, are only readable and usable thanks to some fairly rigid formatting rules. Indeed, when closely examined, the 3 X 5 card seems to communicate as much implicitly by positioning of data, and other cues, as it does explicitly in the characters themselves. Knapp shows clearly how this implicit information must be made explicit during any move to a computerized system.

He discusses the single most obvious characteristic of a bibliographic record, and the data elements in that record, their variable length, and how this impacts system design. The following is a lengthy quote:

> "In the past, format design for data processing has favored the fixed field format because it has advantages in computer processing: fewer instructions need to be written to manipulate the data and processing time on the computer is shorter, and, therefore, more economical. As noted earlier, bibliographic data are not readily adaptable to fixed-field formatting because the data elements in this kind of record have unpredictable lengths.... Of necessity, the formatting of machine-readable cataloging data requires the extensive use of variable fields and the ability to handle records with no prescribed maximum length. Techniques are being developed to deal efficiently with the complexities of bibliographic data in machine processing. Fixed fields, however, can play a role in a format for bibliographic data which could help make the machine-readable record a more powerful tool. Fixed fields may be used:

(1) To make explicit in the machine certain information which is usually implicit to the human, e.g. the country of publication or language of the work, both of which may be expressed in some form of fixed-length code.

(2) To show important characteristics which apply to the whole record but which are not necessarily described by any particular data element, e.g. the work is a government publication.

(3) To make information already carried in a variable field more readily accessible by coding it in a fixed field, e.g. retrieval of records by date of publication may be a common operation and efficiency of retrieval may be worth a redundancy in the record by carrying the information in both the variable and fixed fields.

(4) To augment the catalog record with useful information not usually found on a catalog card, e.g. an indication to show that the work cataloged has an index."

Numerous other considerations which are distinctly system-oriented rather than programming-oriented are treated by Knapp. Perhaps the best single example is Project MARC's decision to express all data characters as a full byte (that's B-Y-T-E, or a sequence of binary digits handled by the computer as a unit) in the MARC II communications format, so that these data may be used on a wide range of computers. For LC's own internal processing operations, a format variation is used which is more efficient on the particular computer installed at the Library. The local format is then translated into the communications format in order to perform the system's very important distribution function.

Project Intrex at MIT provides another fine example of some excellent system design work which has been sensitive to the particular structure of

bibliographic data.

Alan Benenfeld, in a report entitled Generation and Encoding of the Project Intrex Augmented Catalog Data Base (Ref. 6), describes several features which break new ground and which will have to be considered in any extensive future work.

One of these is a Transfer Code:

1. Transfer Code

"A transfer code is used whenever information required for a given catalog record is contained in another record. This is especially useful for relating analytics to their respective whole works. For example, library location and full citation information recorded about an entire conference proceedings need not be recorded again on the separate records for the individual conference papers. A transfer code is used; it contains only the number of the record referred to, which in this case is the record number for the entire conference."

Another intriguing and unusual feature is their User Comments Field, which, years back, when a few people toyed with it as a strictly theoretical concept, I can remember being called "Hypertext."

As Benenfeld describes it:

2. User Comments Field

"Comments will be sought from users on any aspect of this computer catalog, including the indexing, the records, and the documents these represent. These comments will be specially stored and periodically printed out for verification and editing. Comments falling within the sphere of a specific field in a given record will be entered into that field directly. Those comments expressing a value judgment on a document, or pertaining in general to a record will be entered in

field 85 [i.e. a special field]. Comments will be signed, that is, attributed to their source." (p. 24-25)

The Project Intrex work, which is based on a fully "analytical" computer record, does one of the very best jobs I have seen of pointing out just what you gain and lose between a traditional, paragraphed, or run-on record and an analytical record:

"In an analytical record, statements normally found in the body of the descriptive entry of traditional records are broken into component parts, and data of the same kind are listed as a repeating data group. While it is possible to reconstitute traditional statements from listings in an analytical record, this would be inefficient if system output is primarily oriented to providing traditionally formatted printed records. In Intrex, system output is display oriented and the analytically structured record gives added versatility in optimizing displays of bibliographic data.

"Still, if full statements are to be generated from an analytical record, the wording and order may not necessarily be the same as appears on a document title page or in a traditional record. These discrepancies are not considered serious for Intrex because the essential value (content or argument) of each element is retained in the analytical record, and because a document title page can be consulted by display through the Intrex text access system." (p. 25)

In other words, Intrex realizes that once you have pushed Humpty Dumpty off the wall, you can never really get him back together again the way he originally was. Once you have "unitized" a bibliographic record you must forsake knowledge of some of the spatial and contextual relationships that held for the data as it existed in place on the original document. This simple, seemingly trivial, fact has caused much more than its share of problems in library mechanization projects.

C. User Requirements

Moving up the scale from programming problems through systems problems, and beyond, we inevitably run into the poor user's problems. I have neglected him up to now because I have chosen to emphasize the pressures of the data itself rather than the user pressures. Actually, of course, in real life, the priorities should be somewhat reversed. The most beautifully designed system won't be worth a plugged nickel unless the user population likes it, uses it, and finds it satisfies their needs.

The several examples which follow are modular in nature and depending on how much time it is desired to save for questions, I can simply leave a few examples off the far end.

1. All Data Fields Searchable

One very clear user need that emerged in the NASA system that I was involved with was that of increasing the number of data fields that may be queried in the general search. Originally we thought solely in terms of subject matter bibliographies and literature searches. A great deal of other data about each document was gathered and stored; however, it was initially felt that the uses to which this data would be put would be for document control purposes, for statistical, administrative, and management reports, and for document request processing activities independent of the subject searching activity. Various other programs, therefore, made use of these fields, but the original search programs were not constructed so as to exploit very much non-subject data. This turned out to be a false assumption. From the very beginning, we were asked to discriminate within such fields as (1) sponsorship, (2) document and title security levels, (3) country of origin, (4) language, (5) copyright status, (6) contract number, (7) presence or non-presence of a microfiche, etc.

There was thus early demonstrated to us the need to intermix our administrative and bibliographic data right along with the subject indexing data. It is surprising how frequently non-subject data can be utilized to improve or narrow down what is basically a subject matter search. In the new search system now being implemented at the NASA Facility, it will be a rare data element in the file that will not be searchable. The lesson is to generalize your search programs, give yourself as much flexibility as you can by making as many data fields searchable as you can.

2. Root Searching

I am not sure at which point in time it came to our attention that it would be useful to be able to search portions of fields rather than always the whole field. It might easily have started with contract numbers. Many contract numbers have a prefix which indicates from which location they were let and monitored. If one wanted to restrict a search to the output of one or more particular centers, the prefix searching capabilities could be used for this purpose. Many such prefixes can be equated to certain assigned areas of technological responsibility within a total program, and it is nearly possible in this way to restrict a search to a program area.

Much the same argument holds for the other fields where this can be used. There are many report number prefixes which constitute useful "Brand Names" in special search situations. In the field of translations, for instance, the report number prefixes NASA-TT-F, FTD, and JPRS, would immediately be recognized by most special librarians and could be put to practical use.

In the case of personal authors, the technique becomes useful in a slightly different way when one is uncertain of the proper initials,

or when the same man may have been entered in more than one form, as in G. Kuiper and G. P. Kuiper. A search on "Kuiper" alone will save writing whatever variants may exist for Gerald P. Kuiper (though it will also, admittedly pick up some extraneous material by whatever other Kuipers may exist.

Within the area of subject index terms, the technique becomes perhaps even more interesting. For instance, a search on the single root "aluminum" would permit the searcher to include not only the metal itself but all the different aluminum compounds whose names would begin with the word "aluminum." The searcher acquires at least a partial ability to specify "narrower" generic levels without benefit of a formal thesaurus. In non-subject areas it generally makes sense to consider only roots that are prefixes, but within the subject area it is interesting to extend the principle and to consider "floating roots". "Floating roots" refer to particular combinations of letters wherever they appear in a word, beginning, middle, or end; prefix, infix, or suffix. Good examples of useful floating roots might be "pneumo" as in pneumatic, or "iode" as in diode, triode, etc., or "organ" as in organometallic, etc.

3. <u>Ability to Express Logical Role of Every Parameter Queried</u>

Let us move next to the area of logic. In any kind of search effort, there is necessarily some logical relationship among the various parameters queried. This relationship may be expressible at the option of the searcher or it may be "built in" to the system. Our basic method for expressing this relationship had always been the familiar Boolean equation. However, as I have indicated, when we began, our attention was somewhat overly centered on subject searching. In that area we gave ourselves complete flexibility in expressing Boolean relations. Whenever we added non-subject

elements to our search, however, we felt smart enough to decide in advance that we would always want an implied "and" relationship between different fields. For example, if we listed several contract numbers we were automatically asking for all items posted to this contract number or this contract number. When we listed a journal announcement category along with a set of index terms we were automatically "anding" or intersecting the two sets by demanding that both be present. Unfortunately it isn't always necessarily the case that you want to search on this basis, and we again found, therefore, that we had not built sufficient flexibility into the system. Under a concept where all parameters of a search are subject to exact logical expression in an equation, all possible combinations of searching are available to the user. It is this kind of full flexibility in our search logic which we found necessary and towards which the Facility is now headed.

4. Term Weights

Some of you may be familiar with systems that make use of term weights (that is, a numeric value assigned to terms) at the time of retrieval. In such systems, weight values are arbitrarily assigned to each term in the search and the output is controlled by specifying that only items having a certain calculated weight, or greater, be retrieved. We got started with weights in connection with a desire to do something that our Boolean logic couldn't do, or at least as a matter of practicality couldn't do. We would have a group of terms and we wanted to demand that retrieved items be indexed by at least a certain number of these terms, but permitting any combination. This is the same as the "percent matching" technique one so often sees in SDI systems. Boolean logic can't handle it efficiently and so we turned to weights.

No sooner had we introduced weights into our system than several by-products, which we had not originally foreseen, became available. For instance, it is apparent that document weight becomes a way of ranking search output in order of relevance. Probably the first use that weights were put to was not to limit the output, but to arrange it or rank it either for the user or the analyst or perhaps both. This becomes extremely valuable in an environment where search output receives a human edit before it is released. Arbitrary weight levels can be set by the analyst above which relevance to the question is assumed and below which editorial efforts are concentrated.

It next occurred to us that the weighting technique could be made to achieve exactly the same results as a Boolean equation. Cleverly assigned weights could, in a sense, simulate such an equation. Any logical equation can be so converted, though for some equations the process is more cumbersome than for others. This relationship between the two search specification systems had apparently not previously been specifically realized in document retrieval efforts, where they were usually referred to as disparate entities.

We found that there were basically two advantages to having achieved this realization: (1) some searches, especially "percent matching" types are more easily and rapidly expressed by the analyst in terms of weights, and (2) some machines, and the 1410 is one, handle arithmetic techniques faster than logical techniques; therefore, if you can express the logic of a search in terms of weights, the search by that technique will prove to be faster on that type of machine.

The lessons to be had from this experience, other than the power of the weighting technique, are, I suppose, that serendipity will reward

the inquiring mind in this field, as in most others, and, again, that it is advantageous to have your computer doing the things it is best at.

5. Search Expansion Vs. Search Contraction

One of the striking things about bibliographic searching, in my opinion, is the fact that a given search may proceed either as a process of gradual expansion or a process of gradual narrowing down. It may approach the desired set from either end of the spectrum. It all depends on what you start with. Sometimes the searcher in his initial retrieval attempt is faced with a quite large set of documents in which he must find the few that treat of the particular aspect of interest. He must narrow his search. At another time the searcher may address a perfectly reasonable query to the computer and get back the reply that no documents satisfy the conditions specified. In this second case, the searcher must in some way relax the constraints he has imposed and allow at least some documents to pass through the sieves he establishes. He must expand his search. Another situation that may commonly exist, of the latter type, is that in which the searcher actually has an excellent example of a relevant document in hand and wants to find as many more like it as he can. Having observed both problems occur in real-life situations, two areas of further research and development suggested themselves. I do not think anyone has done major work in the design and testing of automatic rules or algorithms for broadening or narrowing a given Boolean expression. In other words, after your first try it becomes clear whether you stand all right or whether broadening or narrowing is required. It can be laborious to successively recode a problem. It should be perfectly possible to have the computer handle the recoding programmatically, following algorithm rules previously worked out. Various levels of the same search would probably

be handled simultaneously under such a system. The European Atomic Energy Agency (EURATOM) currently does something of this sort, I believe, in its bibliographic search in connection with estimating Recall Ratios (or the degree to which the search found what was in the file on the subject in question). Analysts at the National Library of Medicine (NLM) generally code, I believe, three successively looser levels of every retrospective search. Dan Wilde, of the University of Connecticut, in some recent papers dealing with the strategy of interactive searches (see Ref. 42) has also ventured into this area. These are the only cases of which I am aware, however,

A related area where I would like to see some work is that of letting the computer build up its own search from the base of a single document. I would like to be able to go to the computer and say, "Document X is exactly what I am interested in. Print me out citations to the 10 documents that most closely resemble Document X." If someone else doesn't pursue this one, I am sure Leasco eventually will.

IV. Conclusion

In closing, I would merely like to reiterate that this is a very exciting and critical period for library automation, for the standardization of machine input records, and for the design of retrieval systems dealing with these records. Librarians and information scientists are at the brink of a new era. Decisions being made today will literally affect the way we and our children will have access to our bibliographic heritage for decades to come.*

*Relate this to the decisions made in the profession around the turn of the century to go to the unit card and the way these card catalogs are now inextricably related in the public's mind to library service.

APPENDIX A:

Justification for This Topic at an MIS Conference

1. Relation between document retrieval systems and management information systems, in context of information systems in general.
2. Relation between technical information and management processes.

Some of you might well ask what is such a topic doing at a conference where the basic subject is "Management Information Systems". I posed this question myself when first asked to participate here. My personal experience has been almost entirely in library and document handling applications, and I expressed some anxiety about the appropriateness, within the conference framework, as I understood it, of what I felt was perhaps too limited an outlook. I was reassured by my hosts, however, and so here I am and I won't hesitate to implicate them a little if you find the direction I take smacks more of the library than you think it should.

Nevertheless, when I sat down to write, I thought I had better spend at least some time drawing connectives between "Management Information Systems" and the particular kind of information system I wanted to discuss. I wasn't even sure whether the literature of MIS's considered document retrieval systems. I felt that I couldn't proceed into the subject until I had developed some kind of a rationale for wandering, as some might think, so far afield.

At this point I followed the time-honored approach of going to the literature (the "collective wisdom" of the profession) to see if this bridge hadn't already been constructed for me as part of a larger conceptual framework. I was not eminently successful in this search. My general impression is that the field of management information systems does not seem to be at the appr\ ate stage in the development of its basic theory. I did find a few

items, however, which I would like to offer you here by way of justification.

A. "Information Technology: Relationship to Management Decision Models." By Ezra Glaser. (Ref. 17)

I found a little paper by Ezra Glaser entitled "Information Technology: Relationship to Management Decision Models", in which he attempted to classify information systems according to the "complexity" of the information dealt with. At one end of this spectrum was placed "hard" data such as the boiling point of water, which was characterized as "handbook information." At the other end of the spectrum was placed management information systems. One of the most interesting properties of this array of information systems was that at the "hard" end you knew when you had found your answer, whereas at the "soft" end there was an infinite amount and variety of information that the manager might want to have in order to make decisions, and where, in principle, the manager, never achieving omniscience, never had all the information desired. Somewhere in between these two extremes presumably lay library or document handling systems.

B. Information Storage and Retrieval, A State-of-the-Art Report. By Lawrence Berul. (Ref. 7)

Along these same lines, but providing much more elaboration, I found a report I'd had on my shelf for a long time, by Lawrence Berul, entitled Information Storage and Retrieval, A State-of-the-Art Report. This report ties together and synthesizes various strands that had been appearing in the information science literature for years. (See for example F. Jonker, "The Descriptive Continuum", Ref. 24). It does at least two things which I would like to touch on here as part of this preamble. It defines a "communications continuum" which, while it doesn't specifically refer to management information systems, provides an effective framework in which they may be related to all forms of information transfer. It also provides a classification of information

systems, or rather several classifications, which very neatly demonstrates that point of view is everything in developing such a classification, and that any number of schemes are equally valid.

1. Communications Continuum

The "communications continuum" is built on two axes. The horizontal dimension is defined as the "amount of feedback dimension". The best example of high feedback in a communication process is perhaps a two-person conversation with the dialogue providing a direct two-way linkage over which messages are sent. There is real-time stimulus-response situation, on line as it were, with remarks causing remarks, and the behavior of the two participants becoming concerted, cooperative, and directed toward some objective. To quote Mr. Berul: "Newspapers, magazines, and journals provide greater opportunity for communication between the originator and recipient of information as compared to history, archaeology, and cosmology. However, the feedback derived from such a communications link as the letters to the editor of a newspaper or magazine is still several orders of magnitude lower than the feedback provided by person-to-person conversation. The presence or absence of this type of feedback capability is an important consideration in the design of information systems, which are aimed at improving the process of communication. For example, one design consideration is whether the user should be able to conduct a dialogue with a retrieval system either directly with the machine or through an intermediary." (p. 2-2)

The vertical dimension of the so-called "communications continuum" depicts the degree of abstractness of the information being communicated. This refers to the amount of abstract thought necessary to work with the information involved. The low end of this dimension cites such information as logarithmic tables which theoretically do not require abstract thought

because of the lack of ambiguity in ascertaining their meaning. The upper end of this dimension lists music, art, humor, and poetry, ostensibly media creating communication difficulties and uncertainties and demanding considerable abstract thought.

The relevance of this coordinate system to management information systems work seems to be to lie in the realization that any given formal management information system will occupy a different area, a different "zone of retrievability", within the practical limits established by the form of the data being treated. There is no one area of such a chart alloted for MIS's; it depends on their scope and attempted span of control. It is also seen that systems engaged in processing bibliographic data will, in general, occupy a much more limited area of such a chart than will MIS's.

2. Classification of Information Systems.

Moving over to Berul's classification of information systems, we observe that this organization also permits us to better relate management information systems to document retrieval systems. For example, in the scheme having as its organizing principle the "end use" of the output, the information provided by an MIS is almost always intended as an action generator or for monitoring and control, whereas a document system tends to hit more strongly the categories of reference, survey, verification of evidence, etc. It is possible to play this game across all ten classifications provided by Mr. Berul.

C. "Technical Information and Decision-Making".
By Harold Lanier (Ref. 29)

Perhaps, however, in seeking a textbook justification for my dealing with document handling systems at this conference, I am looking for trouble where none really exists. The fact is that it is quite common in the literature to speak of scientific and technical information without regard for the form in which it is embodied or the uses to which it is put. We are all familiar with

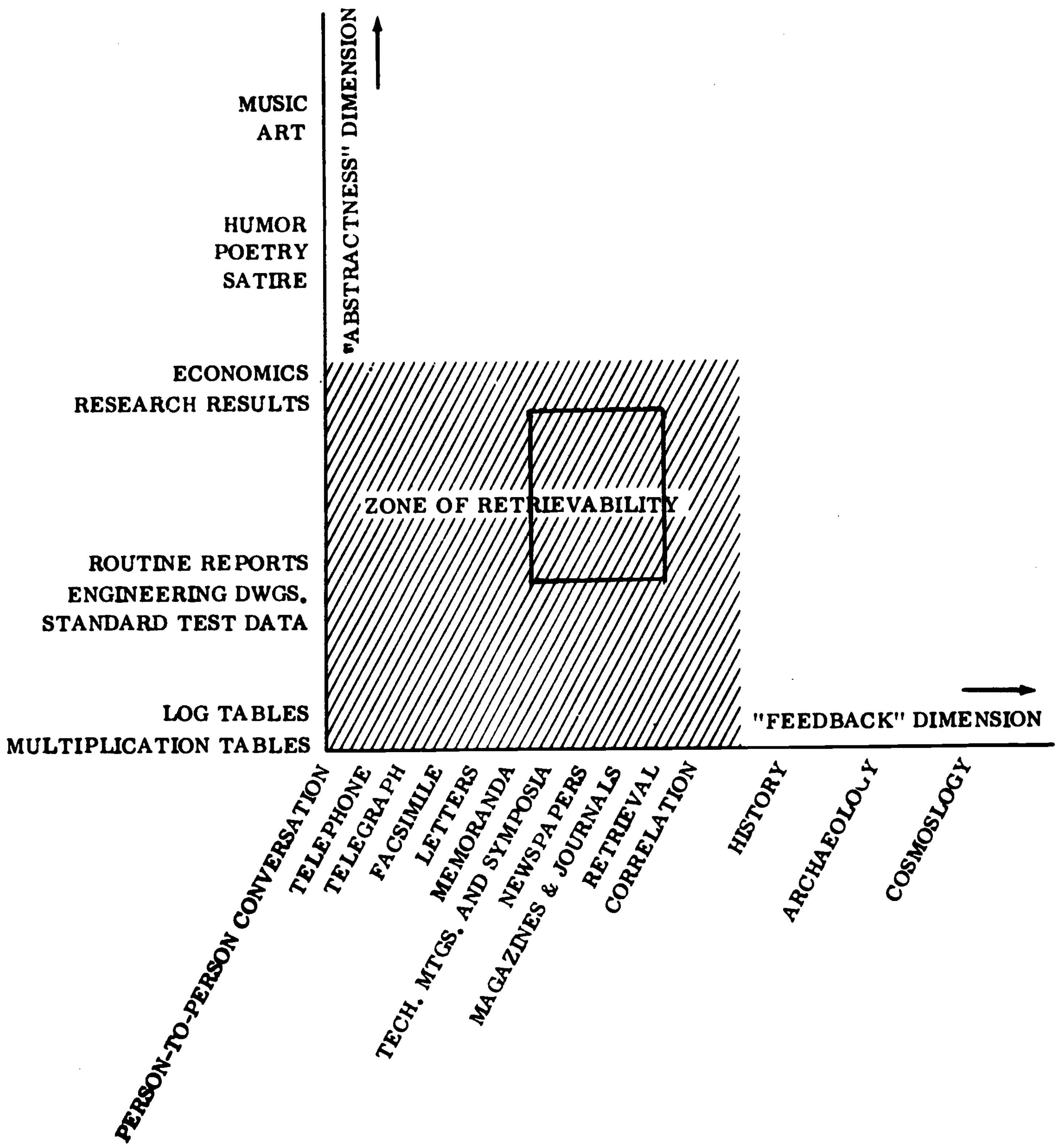
discussions of the "information explosion"; these generally make their argument by citing frightening increases for the volume of documentary material. Likewise, it is almost axiomatic that many of the people who use technical documentary information are in various levels of management and program-planning.

Harold Lanier, in an article entitled "Technical Information and Decision-Making" identifies five classes of information-users, ranging from the individual engineer or scientist to planning groups guiding national programs. He estimates that in the typical industrial organization between one-half and three-quarters of all professional people are engaged at least part time in some "management" facet of the program rather than direct engineering and scientific work. Taking another cut at it, he identifies several progressive stages of information requirements ranging from the isolated scientific fact for solving the particular problem, to related facts, to the rate of acquisition of information, and, finally, trends in the rate of acquisition.

D. NASA as an Adaptive Organization.
by James Webb. (Ref. 41)

This intimate connection between documentary records containing scientific and technical information, and management processes, has perhaps nowhere been more dramatically exemplified than in the National Aeronautics and Space Administration. James Webb, former NASA Administrator, stated in his 1968 Harvard lecture on technological change and management that, "The essence of the job NASA has done is not that a new body of knowledge and technology has been brought into being. Most of the basic knowledge and basic technology was already at hand. The essence of our job has been that of organizing and managing the use of available knowledge and technology in a purposeful and effective way." (p. 23).

With this quotation from Webb, then, I will leave behind the question of justification and attempt to tell you about some of the specific ways in my experience in which the form and nature of bibliographical data, and typical associated data, have affected the design of systems for their control.



REPRINTED FROM:

L. Berul INFORMATION STORAGE AND RETRIEVAL, A STATE OF THE ART REPORT (Ref. 7)

TABLE 2-1. METHODS OF CLASSIFYING INFORMATION SYSTEMS

Class of Information	Subject	Type of Organization	User Job Description	Categories of Use	Mode of Use	Performance Characteristics	Form of Information	Output Media	Type of Information Stored
Concepts	Law	Stock Brokers	Stock Broker	Reference	Current Awareness	Completeness	Items	Teletype	Inventory Data
Cost and Funding	Chemistry	Pharmaceutical Companies	Lawyer	Survey	Retrospective	Relevance	Reports	TV Display	Usage Data
Design Techniques	Medicine	Law Publishers	Doctor	Selection		Specificity	Journal Articles	Microforms	Demand Data
Experimental Processes	Biology	Hospitals	Purchasing Agent	Monitor and Control		Timeliness	Memoranda	Opaque	Engineering Drawings
Math Aids and Formulae	Electronics	R&D Organizations	Salesman	Verification of Exist			Letters	Transparent	Specifications
Performance Characteristics	Supply	Insurance Companies	Personnel Director	Collection			Test Results	Roll	Standards
Production Processes and Procedures	Pharmaceuticals	Transportation Companies	Contract Admin.	Dissem. for Info.			Drawings	35 mm	Test Results
Raw Data	Mathematics	Rate Bureaus	Comptroller	Dissem. for Action			Reference Tools	16 mm	Failure Data
Specifications	Transportation	Police Departments	Intelligence Agent	Dissem. for Future Reference			Abstract Journals	Chip	Maintenance Data
Technical Status	Production	Credit Bureaus	Scientist or Engineer				Indexes	Microfiche	Equipment Population Data
Test Processes and Procedures	Personnel and Training	Airlines	Tech. Evaluation				Encyclopedia	Printed Page	Parts Characteristics
Competitive Intelligence	Management	Banks	Tech. Admin.				Handbooks	Facsimile	Interchangeability Data
Market Intelligence	Metallurgy	Employment Agencies	Research				Information Centers	Computer Printer	Weather Data
	Mathematics	Intelligence Agencies	Expl. Devel.				Bibliographies	Book Form	Maps
	Aircraft and Flight Equip.	Corporations	Adv. Devel.				Correlations	Card Form	Cloud Photography
	Aircraft Instruments	County Recorders	Eng. Devel.				State-of-the-Art Reports		Crop Forecasts
	Aircraft Design	Personnel Departments	Oper. Syst. Dev.				Fact Retrieval Service		Medical Records
	Aircraft Structures	Large Law Firms	Reliability, etc.				Handbooks		Criminal Records
	Flight Operating Problems	Large Consulting Companies							Title Records
	Flight Safety								Patents
	Gliders								

REPRINTED FROM:

L. Berul <u>INFORMATION STORAGE AND RETRIEVAL, A STATE OF THE ART REPORT</u> (Ref. 7)

REFERENCES

1. Sally Alanen
   A Library of Subroutines for Bibliographic Data Processing. n.d. 5 p. Yale University Library.

2. U.S. Atomic Energy Commission, Division of Technical Information.
   Descriptive Cataloging Guide. 2d revision. Oak Ridge, Tennessee, July 1966. 79 p. (TID-4577, rev. 2)

3. Henriette Avram, John F. Knapp, and Lucia J. Rather.
   The MARC II Format -- A Communications Format for Bibliographic Data. Washington, Library of Congress, 1968, 167 p.

4. Henriette D. Avram, Kay D. Guiles, and Guthrie T. Meade.
   "Fields of Information on Library of Congress Catalog Cards: Analysis of a Random Sample, 1950-1964." The Library Quarterly, 37: 180-192 (April 1967)

5. Henriette D. Avram and Julius R. Droz.
   "MARC II and Cobol." Journal of Library Automation, V. 1, No. 4, December 1968, p. 261-272.

6. Alan R. Benenfeld.
   Generation and Encoding of the Project Intrex Augmented Catalog Data Base. August 1968, MIT. 54 p. (ESL-R-360). NSF Grant. NSF-C472.

7. Lawrence Berul.
   Information Storage and Retrieval, A State-of-the-Art Report. Auerbach Corp., Philadelphia, Pennsylvania, September 14, 1964. 249 p.

8. Arthur R. Blum
   Data Element Standardization as an Aid in the Transfer of Physics Information. Proceedings of the American Society for Information Science, Annual Meeting, Columbus, Ohio, October 20-24, 1968, p. 117-119.

9. W. T. Brandhorst and Philip F. Eckert.
   Guide to the Processing, Storage, and Retrieval of Bibliographic Information at the NASA Scientific and Technical Information Facility. Prepared under contract NASw-1315 by Documentation Incorporated for NASA. Washington, D.C., June 1966. 149 p. (NASA-CR-62033)

10. Lawrence F. Buckland.
    The Recording of Library of Congress Bibliographical Data in Machine Form. A report prepared for the Council on Library Resources, Inc., by Inforonics, Inc. February 1965, (Revised). 54 p.

11. Kelley L. Cartwright and Ralph M. Shoffner.
    Catalogs in Book Form. [Berkeley] Institute of Library Research University of California, 1967. 1 V. (various pages)

12. Leonard Cohan, ed.
Directory of Computerized Information in Science and Technology. Science Associates/International, Inc., 1968. loose-leaf. (The International Information Network Services)

13. Committee on Scientific and Technical Information (COSATI)
Bibliographic Data Elements for Machine Tapes for Technical Reports. Report of Panel 1, Operational Techniques and Systems. December 1966. 5 p. (PI-66-RI)

14. Nigel M. Cox and Michael W. Grose, Eds.
Organization and Handling of Bibliographic Records by Computer. 1967, Archon Books. 192 p.

15. Ann T. Curran and Henriette D. Avram.
The Identification of Data Elements in Bibliographic Records. Final Report of the Special Project on Data Elements for the Subcommittee on Machine Input Records (SC-2) of the Sectional Committee on Library Work and Documentation (Z-39) of the United States of America Standards Institute. May 1967. 128 p.

16. C. S. Draper.
"Relationship Between Availability of Data and Significant Technology Advances." _In_ Ralph I. Cole, ed. _Data/Information Availability._ Washington, Thompson Book Co., 1966. p. 43-52.

17. Ezra Glaser
"Information Technology: Relationship to Management Decision Models." _In_ John Houkes, ed., _Management Information Systems and the Information Specialist._ Lafayette, Indiana, 1966. p. 26-35.

18. Marjorie Griffin.
Design Decisions for an Integrated Library System. Los Gatos, California. IBM Advanced Systems Development Division Lab., October 15, 1965.

19. C. D. Gull
"How Will Electronic Information Systems Affect Cataloging Rules." _Library Resources and Technical Services._ n. 5, p. 135-139. Spring 1961.

20. Theodore Hines and Jessica L. Harris
Computer Filing of Index, Bibliographic, and Catalog Entries. Newark, New Jersey, Bro-Dart Foundation, 1966. 128 p.

21. J. M. Houkes, ed.
Management Information Systems and the Information Specialist. Proceedings of a symposium held at Purdue University, July 12-13, 1965. Lafayette, Indiana. 1966. 138 p.

22. International Business Machines Corp., Federal Systems Division.
Report of a Pilot Project for Converting the Pre-1962 National Union Catalog to a Machine Readable Record. Rockville, Maryland, July 1, 1965.

23. International Organization for Standardization.
Bibliographical References; Essential Elements. [Prepared by Technical Committee 130/TC 46] First ed. [Geneva] 1958. 7 p. (ISO Recommendation R77).

24. Frederick Jonker.
"The Descriptive Continuum: A 'Generalized' Theory of Indexing." In Proceedings of the International Conference on Scientific Information, Washington, D.C. November 16-21, 1958. p. 1291-1311.

25. Frederick A. Kilgour
"Symbol-Manipulative Programming for Bibliographic Data Processing on Small Computers." College and Research Libraries, v. 27, no. 2, March 1966, p. 95-98.

26. Frederick A. Kilgour
"Library Catalogue Production on Small Computers." American Documentation, v. 17, no. 3, July 1966. p. 124-131.

27. John F. Knapp
"Design Considerations for the MARC Magnetic Tape Formats." Library Resources and Technical Services, Summer 1968, v. 12, no. 3, p. 275-285.

28. F. H. Lang
The Identification of Input Data in the Library Field. Translated by S. E. Firth. [White Plains, New York, IBM Corp.] February 1967. 30 p.

29. Harold F. Lanier
"Technical Information and Decision-Making." In Ralph I. Cole, ed. Data/Information Availability. Washington, Thompson Book Company, 1966, p. 55-67.

30. Library of Congress, Information Systems Office
Subscriber's Guide to the MARC Distribution Service. Specifications for Magnetic Tapes Containing Monographic Catalog Records in the MARC II format. Washington, August 1968. 68 p.

31. Library of Congress, Information Systems Office.
A Preliminary Report on the MARC (MAchine-Readable Cataloging) Pilot Project. Washington, October 1966. 101 p.

32. Library of Congress, Information Systems Office.
Project MARC, An Experiment in Automating Library of Congress Catalog Data. Washington, June 1967. 17 p.

33. Charles T. Meadow
The Analysis of Information Systems. A Programmer's Introduction to Information Retrieval. New York, Wiley, 1967. 301 p.

34. William R. Nugent
"Compression Word Coding Techniques for Information Retrieval." Journal of Library Automation, v. 1, no. 4, December 1968. p. 250-260.

35. William R. Nugent
"The Mechanization of the Filing Rules for the Dictionary Catalogs of the Library of Congress." Library Resources and Technical Services. v. 11, no. 2, Spring 1967. p. 145-166.

36. Stanford L. Optner & Associates
Report on an Integrated Data Processing System for Library Technical Services to the Public Library, City of Los Angeles. Los Angeles, November 1964. v. 1 (various pages)

37. Phyllis A. Richmond
"Note on Updating and Searching Computerized Catalogs." Library Resources and Technical Services. Spring 1966. p. 155-160.

38. Frederick H. Ruecking, Jr.
"Bibliographic Retrieval from Bibliographic Input; The Hypothesis and Construction of a Test." Journal of Library Automation, v. 1, no. 4, December 1968. p. 227-238.

39. G. Salton
"Data Manipulation and Programming Problems in Automatic Information Retrieval." Communications of the ACM, v. 9, no. 3, March 1966. p. 204-210.

40. Wesley Simonton
"The Computerized Catalog: Possible, Feasible, Desirable?" Library Resources and Technical Services, v. 8, no. 4, Fall 1964. p. 399-407.

41. James F. Webb
NASA as an Adaptive Organization. John Diebold Lecture on Technological Change and Management. Harvard University Graduate School of Business Administration, Boston, Massachusetts, September 30, 1968. 49 p.

42. Daniel U. Wilde
Computer-Aided Strategy Design Using Adaptive and Associative Techniques. Proceedings of the American Society for Information Science, Annual Meeting, Columbus, Ohio, October 20-24, 1968, p. 175-178.

43. Daniel U. Wilde
"Iterative Strategy Design". American Documentation, v. 20, no. 1, January 1969. p. 90-91.

44. Francis A. Williams
"Handling Identifiers as Internal Symbols in Language Processors." Communications of the ACM, v. 2, no. 6, June 1959. p. 21-24.